Reduced Reference Perceptual Quality Model with Application to Rate Control for Video-based Point Cloud Compression

Qi Liu, Hui Yuan, *Senior Member, IEEE,* Raouf Hamzaoui, *Senior Member, IEEE,* Honglei Su, Junhui Hou, *Senior Member, IEEE,* and Huan Yang, *Member, IEEE*

***Abstract*—In rate-distortion optimization, the encoder settings are determined by maximizing a reconstruction quality measure subject to a constraint on the bitrate. One of the main challenges of this approach is to define a quality measure that can be computed with low computational cost and which correlates well with the perceptual quality. While several quality measures that fulfil these two criteria have been developed for images and videos, no such one exists for point clouds. We address this limitation for the video-based point cloud compression (V-PCC) standard by proposing a linear perceptual quality model whose variables are the V-PCC geometry and color quantization step sizes and whose coefficients can easily be computed from two features extracted from the original point cloud. Subjective quality tests with 400 compressed point clouds show that the proposed model correlates well with the mean opinion score, outperforming state-of-the-art full reference objective measures in terms of Spearman rank-order and Pearson linear correlation coefficient. Moreover, we show that for the same target bitrate, rate-distortion optimization based on the proposed model offers higher perceptual quality than rate-distortion optimization based on exhaustive search with a point-to-point objective quality metric. Our datasets are publicly available at https://github.com/qdushl/Waterloo-Point-Cloud-Database-2.0.**

***Index Terms*—Point cloud compression, perceptual quality metric, subjective test, feature extraction, rate-distortion optimization.**

I. Introduction

With the rapid development of three-dimensional (3D) data acquisition technologies, point clouds are now readily available and popular. Point clouds have been used in many applications, including virtual reality, augmented reality, immersive communication, architecture, and autonomous driving [1]. A point cloud comprises a set of points with geometric coordinates and associated attributes, such as color, reflectance, and normal vectors. Point clouds can represent objects or scenes and can be static or dynamic. In this paper, we focus on static point clouds representing objects [2].

To represent the surface of an object with high fidelity, a point cloud usually contains millions or even billions of points. This results in a large amount of data that needs to be efficiently stored and transmitted [3]. The Moving Picture Experts Group (MPEG) has been developing two point cloud compression standards: geometry-based point cloud compression (G-PCC) [4] and video-based point cloud compression (V-PCC) [5]. In these two standards, the geometry and color information is compressed in a lossy way [6], which affects the perceived quality of the reconstructed point clouds.

Like image/video quality assessment methods, point cloud quality assessment methods can be classified into three categories: full reference (FR), reduced reference (RR), and no reference (NR) methods. To evaluate the quality of a distorted point cloud, FR methods use the pristine uncompressed point cloud as a reference, while RR methods only require statistical features that are extracted from the reference point cloud. On the other hand, NR methods evaluate the quality of the distorted point cloud in the absence of the reference.

FR objective quality assessment techniques for point clouds can be based on the point-to-point [7], point-to-plane [8] or point-to-mesh [9] distortion metric. The point-to-point metric computes distances between points in the reference and points in the distorted point clouds, but it does not consider the fact that points in a point cloud usually represent surfaces. The point-to-plane metric projects the error vector along the normal direction of the point in the reference point cloud. The point-to-plane error increases with increasing distance to the local plane surface. The point-to-mesh [9] metric requires the construction of 3D meshes and is therefore not suitable for real-time applications. Other metrics include the angular similarity-based FR metric [10] and the local curvature analysis-based FR metric [11]. Both of them suffer from the high complexity of searching for the neighboring points to construct the normal or curvature. In addition, the above objective quality metrics cannot predict the visual quality of point clouds accurately, especially when the distortion is due to compression [12] [13].

In this paper, we propose an RR method to accurately predict the mean opinion score (MOS) of V-PCC compressed point clouds from the quantization step sizes of the geometry

This work was supported in part by the National Natural Science Foundation of China under Grant 61871342; in part by the open project program of state key laboratory of virtual reality technology and systems, Beihang University, under Grant VRLAB2019B03; in part by the Hong Kong Research Grants Council under Grants 9042955 (CityU 11202320), and in part by Shandong Provincial Natural Science Foundation, China under Grants ZR2018PF002. *Corresponding author: Hui Yuan.*

Q. Liu is with the School of Information Science and Engineering, Shandong University, Qingdao 266237, China. (Email: sdqi.liu@gmail.com).

H. Yuan is with the School of Control Science and Engineering, Shandong University, Ji'nan 250061, China. (Email: huiyuan@sdu.edu.cn).

R. Hamzaoui is with the School of Engineering and Sustainable Development, De Montfort University, Leicester, UK. (Email: rhamzaoui@dmu.ac.uk).

H. Su is with the School of Electronic Information, Qingdao University, Qingdao 266071, China. (Email: suhonglei@qdu.edu.cn).

J. Hou is with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong. (Email: jh.hou@cityu.edu.hk).

H. Yang is with the College of Computer Science and Technology, Qingdao University, Qingdao 266071, China. (Email: cathy_huanyang@hotmail.com).

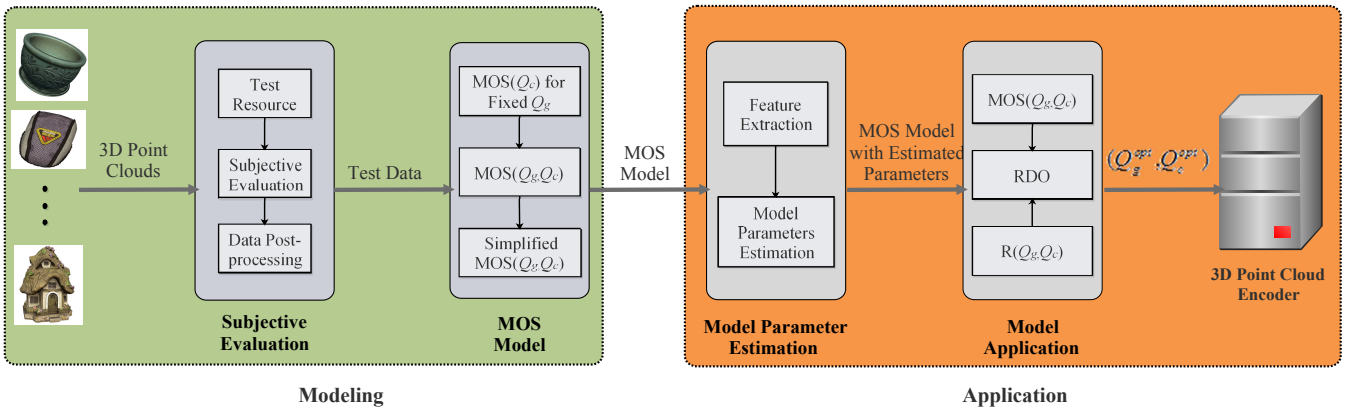


Fig. 1. Compressed point cloud perceptual quality modeling and application to RDO. Q_g and Q_c denote the geometry quantization step size and color quantization step size, respectively, (Q_g^{opt}, Q_c^{opt}) is the optimal geometry and color quantization step size pair, $R(\cdot)$ and $MOS(\cdot)$ are the rate and MOS functions, respectively.

and color video encoders. The proposed model is analytically simple yet effective and can be used for rate-distortion optimized (RDO) rate control (Fig. 1).Note that perceptual quality models [14], [15] for standard video coders cannot be used for V-PCC as these models were developed for different encoder settings (e.g., only one quantization parameter vs. two for V-PCC) and a different distortion type (color distortion vs. both color and geometry distortion for V-PCC). In summary, the main contributions of this paper are as follows:

1) We conduct comprehensive subjective tests to obtain MOSs for point clouds that were compressed with V-PCC using different combinations of geometry and color quantization step sizes. Our subjective tests were carried out for more point clouds and more bitrates than in the literature (400 V-PCC distorted point clouds vs. 180 for the largest existing dataset).
2) We develop a simple yet effective analytical model to predict the MOS from the geometry and color quantization step sizes of the V-PCC encoder. Our model is easier to apply than previous perceptual models as less information is needed to build it (only two variables and three parameters are included). Extensive experimental results show that our model is more accurate than existing FR and RR point cloud quality models.
3) We propose two features to estimate the parameters of our perceptual analytical model with the help of a generalized linear model.
4) We exploit the proposed analytical perceptual distortion model to maximize the subjective quality of the reconstructed point clouds subject to a target bitrate.

The remainder of this paper is organized as follows. Section II briefly reviews related work. Section III describes the subjective study. Section IV proposes a perceptual quality model and validates it using the results of the subjective study. Section V explains how the two features are extracted and how to use them to estimate the model parameters. Section VI proposes a rate control method based on the perceptual quality model introduced in Section IV. Section VII presents an ablation study to demonstrate the robustness and reliability of the proposed perceptual quality model. Finally, Section VIII concludes the paper.

II. Related Work

In this section, we review previous work on subjective quality assessment, objective quality models, and rate-distortion optimized encoding for point clouds.

A. Subjective quality assessment

To develop an accurate perceptual quality model for point clouds, subjective experiments are necessary. In [12], [16]–[18], the geometry distortion was considered, while the distortion due to color was ignored. Torlig *et al.* [19] and Yang *et al.* [20] considered the geometry and color distortions. However, the encoding was not based on V-PCC. Su *et al.* [21], Javaheri *et al.* [22], Alexiou *et al.* [23], and Zerman *et al.* [24] considered the distortion resulting from V-PCC compression. However, the number of distorted point clouds was limited (20 point clouds $\times$ 9 bitrates, 6 point clouds $\times$ 3 bitrates, 9 point clouds $\times$ 5 bitrates, and 2 point clouds $\times$ 4 point counts $\times$ 4 bitrates, respectively).

B. Objective quality models

With the help of the datasets obtained from the subjective experiments, objective quality models have been developed. Yang *et al.* [25] proposed a graph-based objective metric. Although this metric can predict the MOS more accurately than point-to-point, point-to-plane, and point-to-mesh metrics, its complexity is much higher as it requires computationally expensive operations such as resampling of the point cloud and construction of local graphs. Other objective quality models were derived by comparing features extracted from the reference and the distorted point clouds. Alexiou and Ebrahimi [26] extracted and compared a family of statistical local geometry and color features from the reference and the distorted point clouds and used one feature to predict the quality of the distorted point cloud. Meynet *et al.* [27] extracted multiple geometry-based and color-based features and used three of them to construct a distortion model for point clouds. Viola and Cesar [28] extracted 21 features from

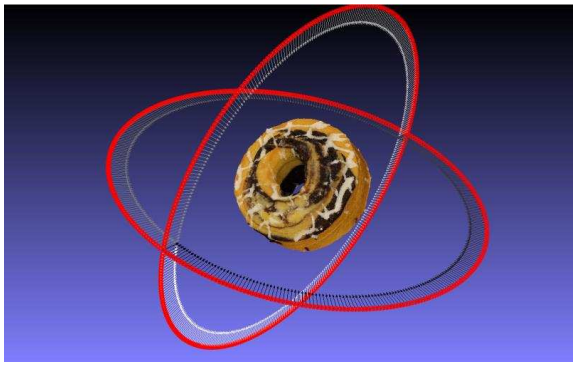


Fig. 2. Generation of pictures from 360 viewpoints of a point cloud.

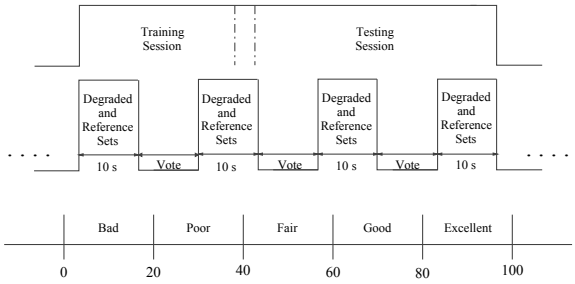


Fig. 3. Schematic diagram of the subjective experiment.

the geometry, luminance, and normal vector information of the point clouds. Then, they predicted the point cloud quality using a linear combination of feature differences where in each difference one feature is extracted from the distorted point cloud and the other feature is extracted from the reference point cloud.

C. Perceptually optimized encoding

In point cloud streaming, the bitrate of the transmitted point cloud must be adapted to the channel bandwidth. To address this requirement, the point cloud can be encoded with different encoder settings. Then, given a target bitrate, the encoder needs to determine the settings that achieve the best perceptual quality. In [29], a model-based rate control technique was developed to efficiently determine the optimal maximum octree level and the JPEG quality factor for the point cloud library-based point cloud compression (PCL-PCC) platform. However, only the color distortion between the original point cloud and the reconstructed point cloud was considered in the bit allocation problem. In [30], a linear combination of the geometry and color quantization step sizes was used to model the point-to-point distortion of V-PCC compressed point clouds. In [31], a coarse-to-fine rate control algorithm was proposed for V-PCC, where the point-to-point distortion metric was also adopted. However, in all these methods, the model for the distortion of the reconstructed point cloud was not based on perceptual quality. Analytical models for the distortion were also used for rate control in image, video, and 3D video communication (e.g., [14], [15], [32], [33]). However, these models cannot be directly applied to V-PCC because both its encoder parameters and the nature and characteristics of point clouds are different.

III. Subjective Quality Assessment

A. Subjective test dataset

It is hard for an observer to perceive the quality degradation of a point cloud with intrinsic distortion [16]. As the number of high-quality point clouds made available by MPEG is limited, for our subjective evaluation, we selected the following 16 point clouds from the Waterloo point cloud (WPC) dataset[1] [21]: Bag (1267845 points), $Banana$ (807184 points), $Biscuits$ (952579 points), $Cake$ (2486566 points), $Cauliflower$ (1936627 points), $Flowerpot$ (2407154 points), $House$ (1568490 points), $Litchi$ (1039942 points), $Mushroom$ (1144603 points), $Ping-pong\_bat$ (703879 points), $Puer\_tea$ (412009 points), $Pumpkin$ (1340343 points), $Ship$ (684617 points), $Statue$ (1637577 points), $Stone$ (1086453 points), and $Tool\_box$ (1054211 points). These point clouds have various geometric and textural features. To encode the point clouds, we used the MPEG V-PCC platform (V-PCC test model v7 [34]) as it normally performs best [6] among the existing public encoders for both static and dynamic point clouds. For each point cloud, we created 25 degraded versions using five geometry QPs (26, 32, 38, 44, and 50) and five color QPs (26, 32, 38, 44, and 50). The corresponding quantization step sizes range from 12.75 to 204. As a result, we had $16\times5\times5 = 400$ point clouds in the subjective evaluation. To display a point cloud as fully as possible, we generated 180 pictures along the horizontal and vertical directions with a step size of two degrees (Fig. 2). Afterwards, the degraded and the original pictures were concatenated to generate 10-second video sequences with 360 frames.

A total of 30 subjects, consisting of 15 males and 15 females aged between 20 and 35, participated in the subjective evaluation. All subjects had normal or corrected-to-normal vision.

B. Subjective evaluation

The double-stimulus impairment scale (DSIS) methodology [35] was adopted in the subjective evaluation. To expand the rating range and obtain finer distinctions, we used a 100-point continuous scale instead of the traditional five-level scale (Fig. 3). To display the stimuli, a DELL E2417H monitor with an in-plane switching display of 23.8 inches (resolution 1920 $\times$ 1080) was used. Both the original and the distorted videos generated from a point cloud were simultaneously shown to the observer side-by-side (Fig. 4). The observer viewed these videos from a distance equal to twice the screen height and rated them through a customized interface after playback.

At the beginning of each evaluation, a training session was conducted to make the observers familiar with the artifacts in the assessment. The point clouds used for training were different from those used for the evaluation. Therefore, the observers were familiar with the distortion types and the quality levels, but not familiar with the content. The duration of each test for a given subject was about two hours, divided

[1]https://github.com/qdushl/Waterloo-Point-Cloud-Database

Fig. 4. Display of original and distorted videos.

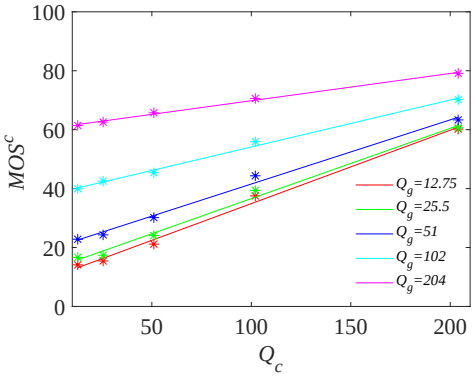


Fig. 5. Relationships between $MOS^c = 100 - MOS$ and Q_c for different Q_gs.

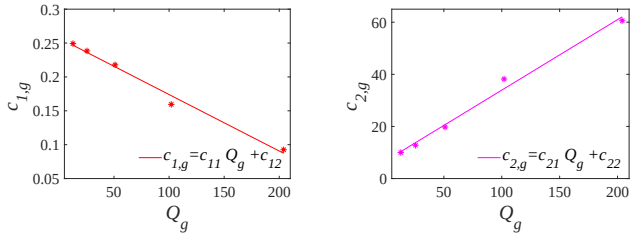
into four sections, with three five-minute breaks in-between to minimize the effect of fatigue.

C. Data post-processing

Since the ratings range from 0 to 100, the scores given by different observers tend to fall in fairly small subranges. Therefore, we converted the subjective scores to Z-scores [36] based on the mean and standard deviation of all the scores of each observer. The Z-score of the m-th point cloud at the j-th degraded level from the i-th viewer is

$$Z_{m,i,j} = \frac{X_{m,i,j} - \mu_{X_i}}{\delta_{X_i}}, \tag{1}$$

where $X_{m,i,j}$ denotes the raw rating, and μ_{X_i} and δ_{X_i} represent the mean and the standard deviation of the ratings of the i-th viewer, respectively. Besides, we adopted the outlier removal technique suggested in [37] to remove outliers. No participants were removed. The obtained Z-scores lied in the range [0, 100]. The MOS of each degraded point cloud was computed as the average of the Z-scores. Using the MOS as the "ground truth", the Pearson linear correlation coefficient (PLCC) and Spearman rank-order correlation coefficient (SRCC) between each viewer's scores and MOS were calculated to verify the performance of individual subjects [21]. Both the mean PLCC and SRCC between each observer scores and the calculated MOS were as high as 0.84, indicating substantial agreement between individual subjects.



(a) (b)

Fig. 6. Relationship between the slope $c_{1,g}$ and intercept $c_{2,g}$ in (2) and Q_g. (a) $c_{1,g}$ vs. Q_g, (b) $c_{2,g}$ vs. Q_g.

TABLE I
ACCURACY OF (2)

Q_g	$c_{1,g}$	$c_{2,g}$	SCC	RMSE	p-value
12.75	0.249	9.986	0.994	1.731	0.000195
25.5	0.238	12.782	0.993	1.785	0.000244
51	0.218	19.765	0.993	1.634	0.000246
102	0.159	38.187	0.994	1.070	0.000176
204	0.093	60.571	0.996	0.525	0.000106

IV. PROPOSED QUALITY METRIC MODEL

To determine the relationship between the perceived quality and the quantization step sizes of the geometry and color, the distorted point clouds with different geometry and color quantization step sizes were rated, as shown in Fig. 5. We can observe that there is a linear relationship between $MOS^c = 100 - MOS$ and the color quantization step Q_c for a fixed geometry quantization step Q_g, that is,

$$MOS^c = c_{1,g}Q_c + c_{2,g}, \tag{2}$$

where $c_{1,g}$ and $c_{2,g}$ are the model parameters. From Table I, we can also see that the squared correlation coefficient (SCC) between MOS^c and Q_c for different Q_gs was larger than or equal to 0.993, while the root mean squared error (RMSE) was smaller than or equal to 1.785. In addition the p-value was smaller than 0.0003 for the null hypothesis that there is no linear relationship between independent variables and dependent variables. As shown in Fig. 6, the relationship between the slope $c_{1,g}$ (respectively the intercept $c_{2,g}$) and Q_g can be represented by the linear models

$$c_{1,g} = c_{11}Q_g + c_{12}, \tag{3}$$

$$c_{2,g} = c_{21}Q_g + c_{22}, \tag{4}$$

where the SCCs of Q_g and $c_{1,g}$, and Q_g and $c_{2,g}$ are 0.988 and 0.990, respectively, and the p-value (for the same null hypothesis as above) is smaller than 0.0006. Accordingly, the quality model can be rewritten as

$$MOS^c = aQ_gQ_c + bQ_g + cQ_c + d, \tag{5}$$

where $a = c_{11}$, $b = c_{21}$, $c = c_{12}$, and $d = c_{22}$. The accuracy of (5) for each point cloud is given in Table II. By further considering the fact that the fitting parameter a is very small (Table II), (5) can be further simplified by removing the impact of $Q_g \cdot Q_c$ on the perceptual quality. This makes the model convex, which is useful in many applications such

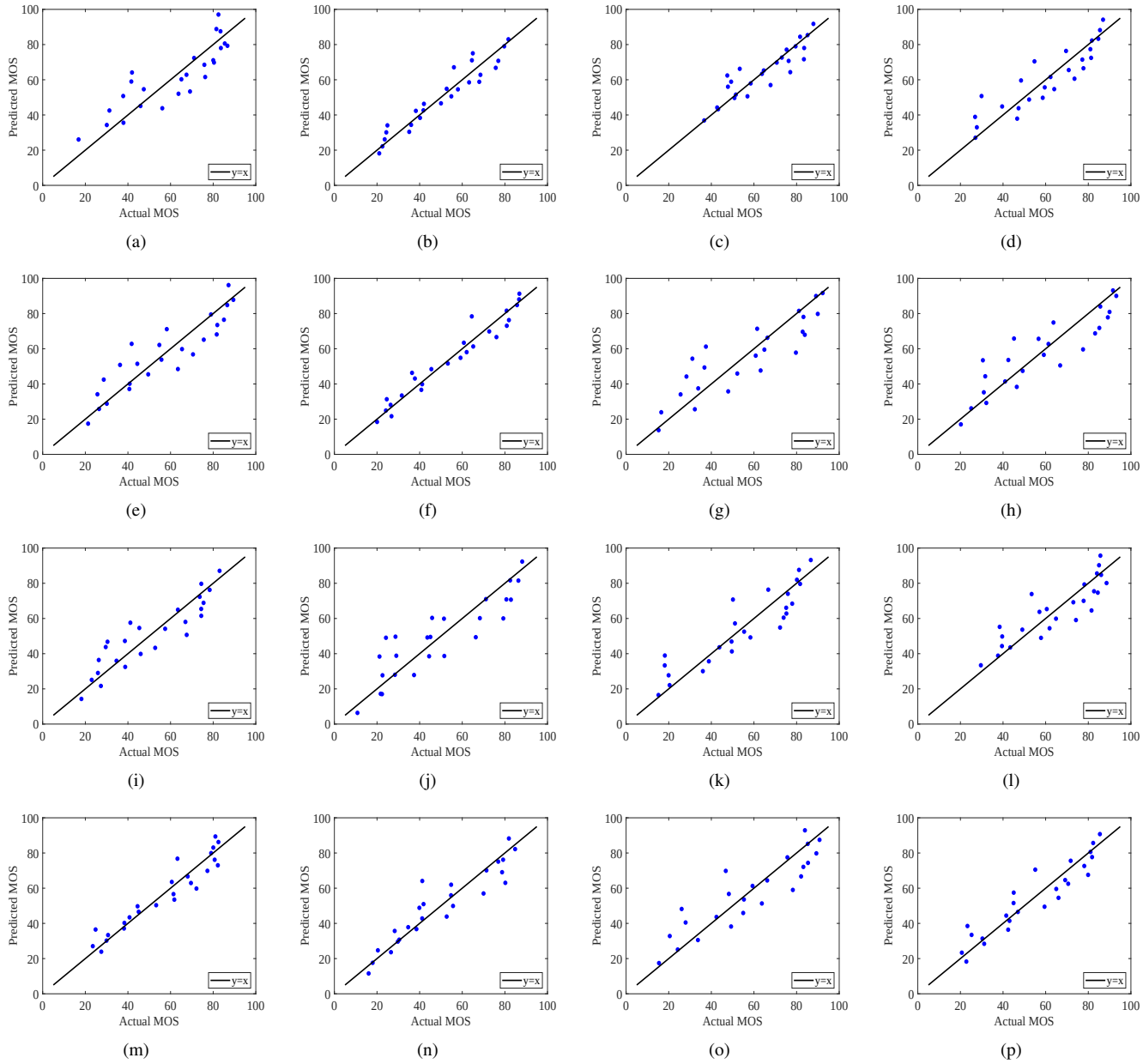


Fig. 7. Accuracy of model (6). (a)-(d): *Bag*, *Banana*, *Biscuits*, and *Cake*, (e)-(h): *Cauliflower*, *Flowerpot*, *House*, and *Litchi*, (i)-(l): *Mushroom*, *Ping-pong_bat*, *Puer_tea*, and *Pumpkin*, (m)-(p): *Ship*, *Statue*, *Stone*, and *Tool_box*.

as rate-distortion optimization. To justify this simplification, we examined the statistical significance of the three terms in (5), i.e., $Q_g \cdot Q_c$, Q_g, and Q_c using a two-way ANOVA test [38]. Table III shows the corresponding F-values, i.e., the MOS^c variations over $Q_g \cdot Q_c$, Q_g, and Q_c (the larger the F-value, the more significant the corresponding term). From Table III, we can see that the statistical significance of $Q_g \cdot Q_c$ is much smaller than that of Q_g and Q_c. Therefore, we further simplified (5) to

$$MOS^c = p_1 Q_g + p_2 Q_c + p_3, \tag{6}$$

where p_1, p_2, and p_3 are model parameters. The model parameters p_1, p_2, and p_3 in (6), the SCCs, and the RMSEs between the actual MOS^cs and the fitted values of all the evaluated point clouds are given in Table IV. We can see that the SCC between the predicted MOS^c and the actual one is between 0.852 and 0.949, with an average value of 0.914, indicating that the derived perceptual quality model is accurate. Fig. 7 illustrates the accuracy of (6). In Section VII-A, we validate (6) on another dataset in an ablation study.

V. Prediction of Model Parameters Using Content Features

Point clouds with rich texture characteristics (e.g., *Cake*) usually have lower MOS^c (corresponding to higher MOS) for the same quantization step sizes. In contrast, point clouds with simple texture characteristics (e.g., *Ping-pong_bat*) have higher MOS^c (corresponding to lower MOS) for the same quantization step sizes. This is because the content has a

TABLE II
ACCURACY OF (5) FOR EACH POINT CLOUD

Point Cloud	a	b	c	d	SCC	RMSE
Bag	-0.0005	0.263	0.223	3.192	0.963	4.317
Banana	-0.0006	0.294	0.127	19.860	0.925	5.663
Biscuits	-0.0006	0.190	0.204	8.293	0.964	3.158
Cake	-0.0008	0.303	0.188	5.519	0.977	3.192
Cauliflower	-0.0010	0.327	0.258	3.389	0.967	4.372
Flowerpot	-0.0005	0.332	0.115	13.016	0.889	8.097
House	-0.0012	0.311	0.361	-3.666	0.981	3.814
Litchi	-0.0012	0.288	0.359	-3.440	0.970	4.536
Mushroom	-0.0010	0.244	0.304	12.295	0.946	5.203
Ping-pong_bat	-0.0014	0.351	0.332	5.463	0.951	5.875
Puer_tea	-0.0009	0.192	0.366	6.488	0.982	3.379
Pumpkin	-0.0007	0.184	0.276	3.242	0.969	3.557
Ship	-0.0006	0.312	0.112	13.296	0.928	5.905
Statue	-0.0007	0.308	0.196	14.527	0.874	8.496
Stone	-0.0010	0.245	0.366	-1.385	0.981	3.588
Tool_box	-0.0008	0.184	0.333	9.886	0.951	5.124

TABLE III
TWO-WAY ANOVA ON MOS^c

Factors	Q_g	Q_c	$Q_g \cdot Q_c$
F-value	226.802	197.838	4.660

TABLE IV
PARAMETERS AND ACCURACY OF THE PERCEPTUAL QUALITY MODEL

Point Cloud	p_1	p_2	p_3	SCC	RMSE	p-value
Bag	0.223	0.183	6.342	0.949	4.954	1.08×10^{-11}
Banana	0.247	0.080	23.601	0.902	6.336	2.45×10^{-4}
Biscuits	0.143	0.156	12.072	0.927	4.387	2.21×10^{-11}
Cake	0.241	0.125	10.489	0.938	5.153	2.34×10^{-8}
Cauliflower	0.246	0.177	9.773	0.916	6.782	7.02×10^{-9}
Flowerpot	0.291	0.075	16.212	0.877	8.339	4.95×10^{-3}
House	0.220	0.269	3.597	0.930	7.059	5.81×10^{-12}
Litchi	0.195	0.266	3.874	0.914	7.488	2.23×10^{-11}
Mushroom	0.164	0.225	18.579	0.890	7.262	3.10×10^{-10}
Ping-pong_bat	0.240	0.221	14.240	0.872	9.243	3.13×10^{-8}
Puer_tea	0.124	0.297	11.921	0.948	5.568	1.00×10^{-14}
Pumpkin	0.131	0.223	7.424	0.939	4.898	1.60×10^{-13}
Ship	0.268	0.068	16.756	0.910	6.438	1.30×10^{-3}
Statue	0.254	0.142	18.777	0.852	9.011	1.64×10^{-5}
Stone	0.170	0.291	4.555	0.945	6.026	5.00×10^{-14}
Tool_box	0.117	0.266	15.152	0.914	6.630	2.01×10^{-12}
Average	-	-	-	**0.914**	**6.598**	4.07×10^{-4}

concealing effect on the coding distortion, which is consistent with the characteristics of the human visual system [39]. Thus, the model parameters in (6) will be highly content dependent. In this section, we propose two features to predict the model parameters efficiently. The perceptual quality of a point cloud depends on both the geometry and color distortion. But the influence of geometry distortion is different from that of the color distortion [22]. By analyzing the local topological and color consistencies, Alexiou and Ebrahimi [26] and Meynet *et al.* [27] reported that color-based features achieve the best performance in predicting the perceptual quality. Accordingly, we extracted two novel texture features (a local feature and a global feature) to predict the model parameters effectively. The local feature represents color fluctuation over geometric distances (CFGD), while the global feature is the color block mean variance (CBMV).

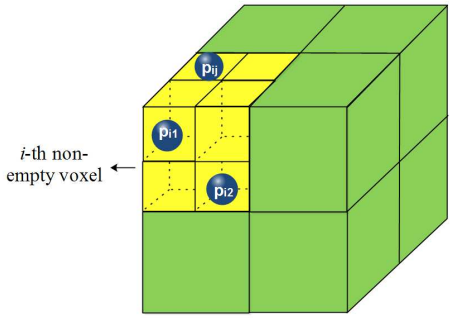


Fig. 8. Voxelized point cloud. The voxel size can be 8^3, 16^3, 32^3, or 64^3

A. Color fluctuation over geometric distances (CFGD)

For a point p_i in the point cloud, we first define $CFGD_i$ as

$$CFGD_i = \frac{1}{N_i} \sum_{p_j \in \mathbb{S}_i} \frac{|C(p_i) - C(p_j)|}{d_{i,j}}, \tag{7}$$

where $C(\cdot)$ denotes the color attribute of a point, $d_{i,j}$ denotes the distance between points p_i and p_j, $\mathbb{S}_i$ is the set of the K nearest neighbors of point p_i, and N_i is the number of points in $\mathbb{S}_i$. For simplicity, we only consider the Y (luminance) component [40] in this paper. Then, CFGD is defined as

$$CFGD = \frac{1}{T} \sum_{i \in \mathbb{P}} CFGD_i, \tag{8}$$

where T is the number of points in the point cloud $\mathbb{P}$.

B. Color block mean variance (CBMV)

The standard deviation is commonly used as a global feature for image/video quality assessment [41] [42] [43]. Similarly, we use it to build a global feature for point clouds. Assuming that the point clouds are voxelized [44] (Fig. 8), CBMV is computed as

$$CBMV = \frac{1}{B} \sum_{i=1}^{B} \sqrt{\frac{1}{D_i} \sum_{j=1}^{D_i} (C(p_{ij}) - \mu_i)^2}, \tag{9}$$

where B denotes the number of non-empty voxels, D_i denotes the number of points in the i-th non-empty voxel, $C(p_{ij})$ is the color of the j-th point in the i-th non-empty voxel, and μ_i is the color mean value of the i-th non-empty voxel.

C. Model parameter estimation

As in [14], we used a generalized linear model (GLM) [45] to predict the model parameters from the extracted two features. Given a training set of M point clouds, let $p_{m,j}$ denote the j-th parameter in (6) of the m-th point cloud ($j = 1, 2, 3; m = 1, 2, ..., M$). Let $f_{m,k}$ denote the value of the k-th feature for the m-th point cloud, where $1 \leq k \leq K$, and K is the number of extracted features (in this paper, $K = 2$).

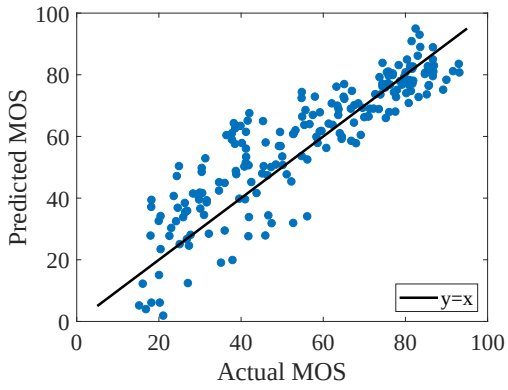


Fig. 9. Scatter plot of the actual MOS and the MOS predicted by the proposed quality model for the test set.

Then, the parameter $p_{m,j}$ is estimated by a generalized linear predictor

$$\hat{p}_{m,j} = h_{j,0} + \sum_{k=1}^{K} f_{m,k} h_{j,k}, \tag{10}$$

where $h_{j,0}$ is a constant weight of the j-th parameter, and $h_{j,k}$ is the weight of the j-th parameter of the k-th feature. The generalized linear predictor can be expressed in vector form $\hat{\mathbf{P}}_\mathbf{m} = \mathbf{F}_\mathbf{m}\mathbf{H}$, where $\hat{\mathbf{P}}_\mathbf{m} = [\hat{p}_{m,1}, \hat{p}_{m,2}, \hat{p}_{m,3}]$, $\mathbf{F_m}$=$[1, f_{m,1}, f_{m,2}]$, and $\mathbf{H}$ is a $3\times(K+1)$ coefficients matrix with elements $h_{j,k}$. The aim is to find a matrix $\mathbf{H}$ that minimizes the prediction error

$$\varepsilon = \sum_{m=1}^{M} \|\hat{\mathbf{P}}_\mathbf{m} - \mathbf{P}_\mathbf{m}\|_2^2, \tag{11}$$

where $\mathbf{P_m} = [p_{m,1}, p_{m,2}, p_{m,3}]$.

In this paper, $\mathbf{H}$ was obtained by minimizing (11) for voxel size 64^3 and a training set of $M = 8$ point clouds that cover a wide range of content characteristics ($Cauliflower$, $Stone$, $House$, $Ship$, $Tool\_box$, $Pumpkin$, $Biscuits$ and $Ping-pong\_bat$). The remaining point clouds, i.e., $Litchi$, $Puer\_tea$, $Flowerpot$, Bag, $Cake$, $Statue$, $Banana$ and $Mushroom$ were used for testing. Minimization of (11) gave the optimal solution

$$\mathbf{H^{opt}} = \begin{bmatrix} 0.1817 & 0.2058 & 18.4528 \\ 0.0034 & -0.0070 & -0.0199 \\ -0.0116 & 0.0292 & -1.5427 \end{bmatrix}. \tag{12}$$

By using $\mathbf{H^{opt}}$ and the extracted feature vector $\mathbf{F_m}$, the model parameter vector $\hat{\mathbf{P}}_\mathbf{m}$ can be calculated directly. Furthermore, based on the estimated model parameters, we can obtain the MOS^c through (6). We use PLCC, SRCC, and RMSE between the actual MOS^cs and the predicted ones to evaluate the accuracy of the proposed model with the estimated parameters. Table V shows that the PLCC and SRCC of the proposed perceptual quality model of the test set are up to 0.9133 and 0.9095, respectively, and the RMSE is as small as 8.9090 (noting that the maximum MOS is 100). Fig. 9, which shows the relationship between the actual MOSs and the estimated ones, illustrates the accuracy of the model.

TABLE V
PERFORMANCE OF THE PERCEPTUAL QUALITY MODEL ON THE TRAINING AND TEST SETS. V_{size} IS THE VOXEL SIZE FOR CBMV

DataSet	V_{size}	PLCC	SRCC	RMSE
Training Set	8^3	0.9291	0.9358	8.1530
	16^3	0.9335	0.9377	7.9047
	32^3	0.9369	0.9402	7.7078
	64^3	0.9377	0.9409	7.6597
Test Set	8^3	**0.8963**	**0.8922**	**9.7016**
	16^3	0.8998	0.8972	9.5457
	32^3	0.9080	0.9053	9.1651
	64^3	**0.9133**	**0.9095**	**8.9090**

Next, we compare the proposed RR quality model to the state-of-the-art. First, we compare our model to another RR model [28] (Table VI). The RR model in [28] calculates the perceptual quality score as a linear combination of 21 feature differences. The results show that the model in [28] was less accurate than ours as its PLCC and SRCC were only 0.3851 and 0.3940, respectively. Next, we compared our model to several representative FR objective metric models: one point-based model [9], three projection-based models [19] [46] [47] [48], one graph-based model [25], and two feature-based models [26] [27]. The point-based method, which is adopted by MPEG, captures the difference between the points in the reference and the tested point cloud. For the projection-based approaches, a point cloud is mapped onto six two-dimensional image planes by orthographic projection. After obtaining the projected image planes, the 2D image quality metrics structural similarity index measure (SSIM) [47], multi-scale structural similarity index measure (MS-SSIM) [48], and visual information fidelity in pixel domain (VIFP) [46] are used to evaluate the quality of the six image projections. Finally, the average image quality of these six projections is mapped to MOS by the best fitting logistic function, and the mapped MOS is taken as the quality of the point cloud. We name these projection-based methods SSIM$_{projection}$, MS-SSIM$_{projection}$, and VIFP$_{projection}$, respectively. For the graph-based method in [25] called GraphSIM, local graphs centered at the key points were used to calculate the similarity between the original and the distorted point cloud. In the two feature-based models (PointSSIM [26] and PCQM [27]) the quality is assessed by comparing the feature difference between the reference and the distorted point clouds. Table VI compares the point-based and projection-based methods. We can see that the point-based $PSNR_Y$ model does not seem to provide enough accuracy. GraphSIM improves the prediction accuracy to some extent; however, it is more complex and requires many parameters to be determined. The projection-based models perform best, with VIFP achieving the best performance compared to PSNR, SSIM and MS-SSIM. Nevertheless, the accuracy of the prediction is only modest compared with their performance on 2D images [21]. Table VI shows that the PLCC of FR quality metrics is in the range 0.3369 to 0.8405. In contrast, the PLCC of the proposed RR quality metric is about 0.9133. In addition to the PLCC, the SRCC of the worst and best FR quality metrics are 0.3470 and 0.8368, respectively, whereas the SRCC of the proposed RR quality

TABLE VI
PERFORMANCE OF POINT CLOUD QUALITY ASSESSMENT MODELS.

Model Type	Model	|PLCC|	|SRCC|
FR	PSNR_Y	0.3900	0.3705
	$\text{SSIM}_{projection}$	0.3369	0.3470
	$\text{MS-SSIM}_{projection}$	0.4734	0.4606
	$\text{VIFP}_{projection}$	0.8405	0.8368
	GraphSIM	0.7590	0.7568
	PCQM	0.6569	0.6727
	PointSSIM	0.4645	0.4847
RR	PCM_{RR}	0.3851	0.3940
	Proposed ($V_{size} = 64^3$)	**0.9133**	**0.9095**

TABLE VII
TARGET BITRATES IN KILOBITS PER MILLION POINTS (*kbpmp*) FOR EACH POINT CLOUD IN THE TEST SET

Point Cloud	$R_{T,1}$	$R_{T,2}$	$R_{T,3}$	$R_{T,4}$
Bag	170	510	1495	2130
Banana	40	120	310	850
Cake	110	170	265	460
Flowerpot	75	135	265	405
Litchi	110	250	565	1200
Mushroom	50	150	220	375
Puer_tea	75	190	640	1525
Statue	55	105	155	200

metric is 0.9095.

VI. APPLICATION TO RATE CONTROL

The developed perceptual quality model can benefit applications involving coding and rate control. In this section, we propose a rate control method for static point clouds. Our method can also be extended to dynamic point clouds as they can be seen as a sequence of successive static point clouds. For a given target bitrate, we aim to find the combination of the geometry QP (corresponding to Q_g) and color QP (corresponding to Q_c) that provides the best perceptual quality. We formulate this rate control problem as a constrained optimization problem where the objective function is the derived perceptual quality model. That is, we consider the problem

$$\begin{aligned} &\min_{(Q_g, Q_c)} MOS^c(Q_g, Q_c) \\ &\text{s.t.} \quad R_g(Q_g) + R_c(Q_g, Q_c) \leq R_T, \end{aligned} \tag{13}$$

where R_g and R_c are the geometry and color bitrate, respectively, and R_T is the overall target bitrate. Based on (6) and the Cauchy-based rate model [30], the rate control problem (13) can be rewritten as

$$\begin{aligned} &\min_{(Q_g, Q_c)} p_1 Q_g + p_2 Q_c + p_3 \\ &\text{s.t.} \quad \gamma_g Q_g^{\theta_g} + \gamma_c Q_c^{\theta_c} \leq R_T, \end{aligned} \tag{14}$$

where p_1, p_2, and p_3 are the parameters of the perceptual quality model, and γ_g, θ_g, γ_c, θ_c are the parameters of the geometry and color rate models.

To solve (14), we first extract CFGD and CBMV from the input point cloud, as described in Section V. Then by using the pre-trained matrix $\mathbf{H}$ in (12), we compute the parameter vector $\hat{\mathbf{P}}_\mathbf{m}$. Next, we encode the point cloud with two pairs of geometry and color quantization parameters to determine the parameters γ_g, θ_g, γ_c, and θ_c. Finally, we use the barrier interior point method in [30] to solve (14).

To validate the proposed perceptual quality model-based rate control algorithm, we compared its performance to that of a point-to-point based exhaustive search algorithm (denoted by $P2P_{ES}$). For $P2P_{ES}$, a point cloud was first encoded by all the tested geometry and color QP pairs ranging from 26 to 50 with step size 6. Then the subset of admissible pairs (pairs for which the bitrate is smaller than or equal to the target bitrate) was determined. Finally, the pair that gave the highest PSNR for the Y component ($PSNR_Y$) was selected from this subset. We focused on the Y component because it plays an important role in visualization and in our perception of objective structure and surface shape [49]. Since the texture complexity varies between the tested point clouds, we set different target bitrates for each point cloud (Table VII). The rate-MOS curves of the proposed algorithm and $P2P_{ES}$ are compared in Fig. 10. The results demonstrate that the proposed rate control algorithm can achieve better rate-MOS performance than $P2P_{ES}$ with much lower complexity. Since the value of $PSNR_Y$ in $P2P_{ES}$ is not consistent with the MOS, the MOS for $P2P_{ES}$ does not necessarily increase with the target bitrate. The proposed algorithm used the proposed RR model to better predict the MOS, and achieved a better subjective quality for a given target bitrate. Finally, Fig. 11 compares the subjective quality of point clouds obtained with the proposed rate control algorithm and $P2P_{ES}$. We can see that a significant improvement in subjective quality can be achieved with the proposed RR model-based rate control algorithm.

VII. ABLATION STUDY

In this section, we first verify the accuracy of the quality model (6) on two other V-PCC subjective datasets. Then, we study feature selection for the prediction of the model parameters and discuss the reliability of the proposed quality model. Finally, we discuss the impact of the $CBMV$ block size on the rate control algorithm.

A. Performance of the Quality Model on Other Datasets

To verify the validity and reliability of the proposed quality model (6), we evaluated it using the M-PCCD [23] and VSENSE VVDB [24] datasets.

The VSENSE VVDB dataset consists of two dynamic point clouds ($Matis$ and $Rafa$) showing a football game. These two dynamic point clouds were sampled at four sampling frequencies and compressed using V-PCC with four quantization parameter combinations [(17,20), (30,35), (37,43), (41,48)]. Accordingly, 32 distorted point clouds were generated. In the subjective tests, the MOS range was [0, 100] (refer to [24] for details). The accuracy of the proposed quality model (6) on the VSENSE VVDB dataset is shown in Table VIII. We can see that the average SCC of the proposed quality model (6) is about 0.97.

The M-PCCD dataset consists of eight static point clouds ($longdress$, $loot$, $soldier$, $the20smaria$, $amphoriskos$,

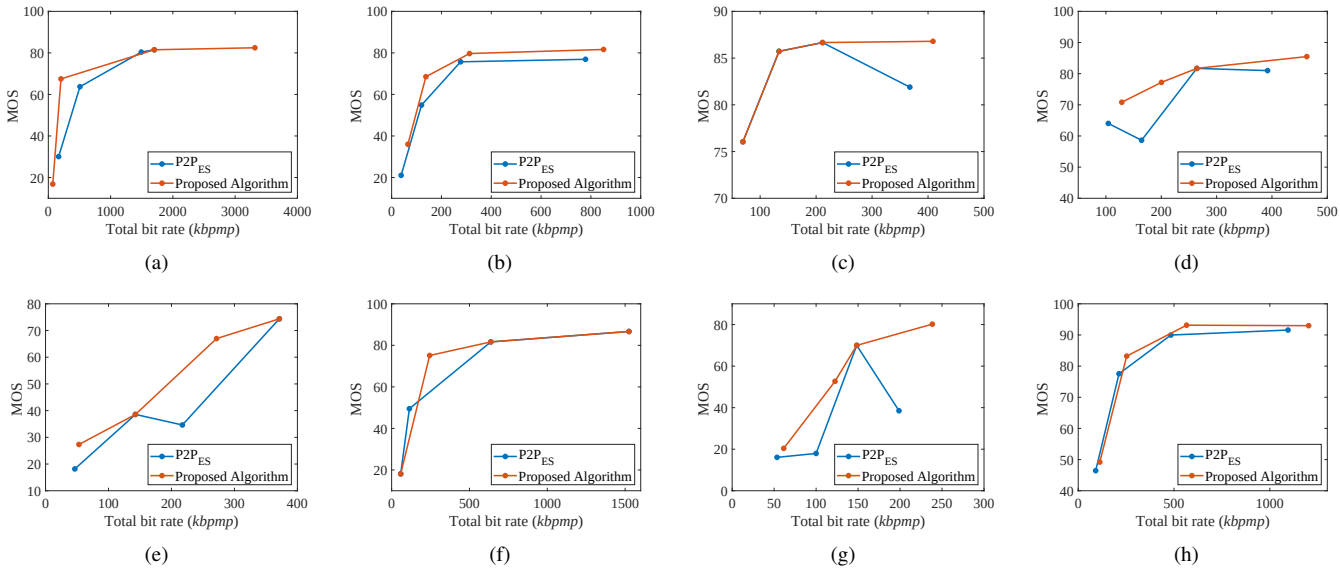


Fig. 10. MOS vs. total bitrate for our rate control algorithm and the point-to-point exhaustive search algorithm ($P2P_{ES}$). (a) *Bag*, (b) *Banana*, (c) *Flowerpot*, (d) *Cake*, (e) *Mushroom*, (f) *Puer_tea*, (g) *Statue*, (h) *Litchi*. The vertical axis shows the average MOS of all the viewers.

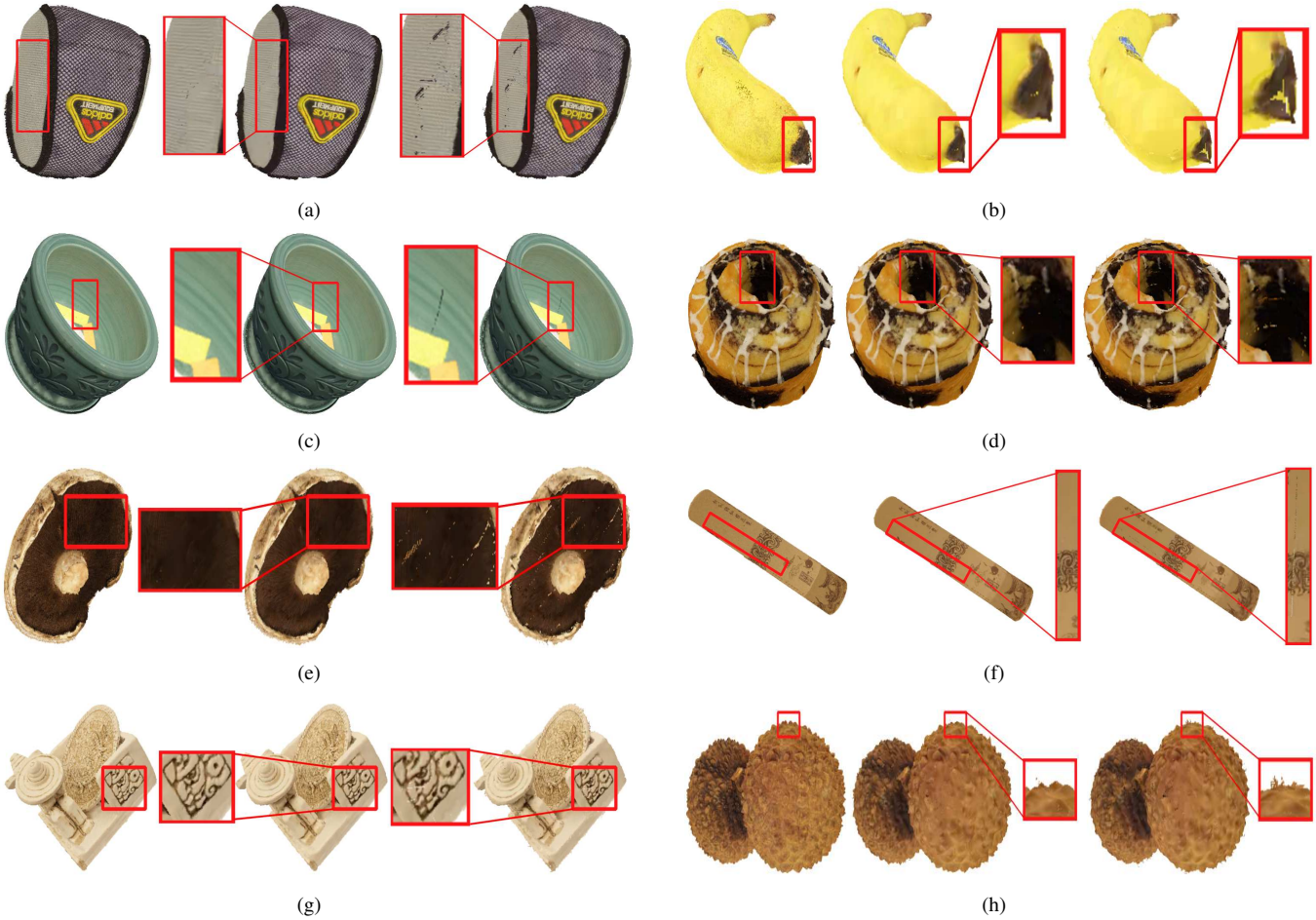


Fig. 11. Perceptual quality comparison between our rate control algorithm and $P2P_{ES}$. Left: original, Centre: proposed, Right: $P2P_{ES}$. (a)subjective quality of *Bag* with a target bitrate of 510 kbpmp, (b)subjective quality of *Banana* with a target bitrate of 85 kbpmp, (c)subjective quality of *Flowerpot* with a target bitrate of 405 kbpmp, (d)subjective quality of *Cake* with a target bitrate of 170 kbpmp, (e) subjective quality of *Mushroom* with a target bitrate of 275 kbpmp, (f)subjective quality of *Puer_tea* with a target bitrate of 190 kbpmp, (g)subjective quality of *Statue* with a target bitrate of 165 kbpmp, (h)subjective quality of *Litchi* with a target bitrate of 110 kbpmp.

TABLE VIII
ACCURACY OF (6) ON THE VSENSE VVDB DATASET.

Point Cloud Sequence	p_1	p_2	p_3	SCC	RMSE
Matis	4.07	-1.36	13.31	0.97	4.77
Rafa	2.97	-0.90	16.71	0.97	4.48
Average				0.97	4.63

TABLE IX
ACCURACY OF (6) ON THE M-PCCD DATASET.

Point Cloud	p_1	p_2	p_3	SCC	RMSE
amphoriskos	-0.15	0.12	0.58	0.99	0.10
biplane	-0.33	0.17	1.50	0.82	0.59
head	0.19	0.08	-0.44	0.99	0.22
longdress	0.29	0.05	-0.40	0.97	0.31
loot	0.07	0.06	1.28	0.99	0.10
romanoillamp	0.02	0.12	0.03	0.99	0.17
soldier	-0.09	0.13	1.01	0.94	0.40
the20smaria	0.15	0.04	0.48	0.89	0.42
Average				0.95	0.29

$biplane$, $head$, and $romanoillamp$) encoded with both V-PCC and G-PCC. For V-PCC, five distortion levels were generated for each point cloud by varying the geometry and color quantization parameters. In the subjective tests, the MOS range was [0,5]. More details about the subjective test can be found in [23]. Using the V-PCC codec, the degradation levels were obtained by modifying the geometry and texture quantization parameters. Finally, 40 distorted point clouds were used for verification. The accuracy of the proposed quality model (6) on the M-PCCD dataset is shown in Table IX. Note that $MOS^c = 5 - MOS$ because a 5-point scale MOS is used in this dataset. We can see that the average SCC of the proposed quality model (6) is about 0.95, confirming its effectiveness.

B. Selection of Feature Sets for the Model Parameter Estimation

In addition to the proposed local and global features (i.e., CFGD and CBMV), we also considered existing features that have proven effective in subjective quality prediction: a geometry one (the curvature-based feature in [26]) and two color ones (the luminance mean feature and the luminance entropy feature in [28]).

The curvature feature is computed as follows [26]. First, the coordinates of the centroid of the points in the point cloud are subtracted from the geometric coordinates of each point to establish a local coordinate system. Then, a Kd-tree is created based on the local coordinate system and the K nearest neighbors of each point are obtained. Next, the curvature of each point is calculated according to the coordinate information of the current point and its neighbors [50]. Finally, the curvature feature ($G_{\delta_{cur}}$) is obtained as the variance of the curvatures of the points in the point cloud.

The two color features are obtained as follows. The color attributes R, G, B are first converted to luma, chroma blue (Cb), and chroma red (Cr) format by using the matrix defined in ITU-R Recommendation BT.709 [51]. Then the mean value of the luminance is calculated to obtain one color feature (C_M). In addition, the entropy of the luminance is calculated as another color feature (C_E) [28]. Details about the extraction of the color features can be found in [52].

To find the smallest set of features that can predict the model parameters accurately, we conducted an ablation study. We first explored the correlation coefficients between the features (i.e., $CFGD$, $CBMV$, $G_{\delta_{cur}}$, C_M, and C_E) and the optimally fitted parameters (i.e., p_1, p_2, and p_3) of all point clouds to select a basic set of features. Then we randomly picked eight point clouds for training and used the remaining ones for testing. Based on the training set, we increased the number of features in the basic set one by one to predict the optimal model parameters. Next, by using the testing set, we computed the correlation coefficients (CCs) between the predicted model parameters stemming from the selected features and the optimally fitted parameters. Finally, according to the CCs, we determined the final feature set to predict the model parameters. Details of the process are given below.

The basic set of features was determined based on the correlation coefficients between the features (i.e., $CFGD$, $CBMV$, $G_{\delta_{cur}}$, C_M, and C_E) and the optimally fitted parameters (i.e., p_1, p_2, and p_3) of all point clouds. Based on the results in Table X, we selected the features with the hightest correlation coefficients, i.e., $CFGD$ and $CBMV$, as basic features. Then we randomly selected eight point clouds as the training set to calculate the optimal weight matrix $\mathbf{H}$ (see Section V-C). The remaining point clouds were used for testing. Next, we tried to increase the number of selected features in the training set. Following the process in Section V-C, the predicted model parameters (expressed by vector $\hat{\mathbf{P}}$) of the point clouds can be calculated by $\hat{\mathbf{P}} = \mathbf{FH}$ based on the testing dataset, where $\hat{\mathbf{P}}$ is the predicted model parameter vector of one point cloud, and $\mathbf{F}$ is the selected feature vector for one point cloud. Once $\hat{\mathbf{P}}$ for the point clouds in the test set is calculated, the corresponding CC with $\mathbf{P}$ can be computed. The above process was repeated three times. Table XI shows the results. We can see that increasing the number of features did not significantly increase the CC and can even decrease it. Therefore, for simplicity, we used only CFGD and CBMV.

C. Reliability of Proposed Method

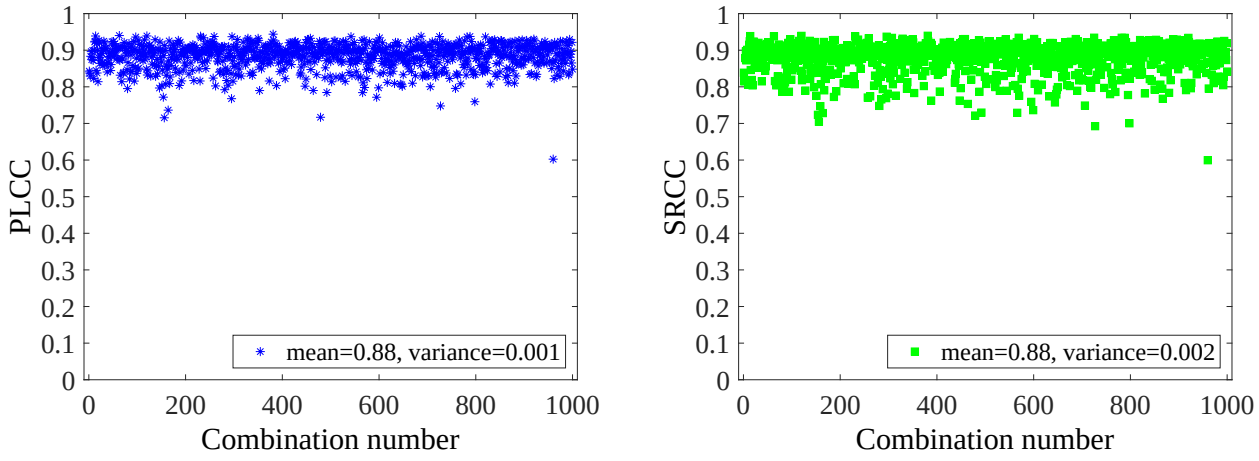
To verify the reliability of the proposed method, we calculated the PLCC and SRCC between the ground truth MOS and predicted MOS for 1000 random combinations of the point cloud, Q_g, and Q_c. The results, shown in Fig. 12, confirmed the accuracy of the proposed model (the average PLCC and the average SRCC were both equal to 0.88).

D. Influence of Block Size on Rate Control

In this part, we further explore the influence of the CBMV block size on the rate control performance. The results are shown in Table XII, where R_T is the target bitrate, QP_G^{M8} and QP_C^{M8} are the geometry and color QP calculated by solving (14) when V_{size} was set to 8. Similarly, (QP_G^{M16}, QP_C^{M16}), (QP_G^{M32}, QP_C^{M32}), (QP_G^{M64}, QP_C^{M64}) are the geometry and color QP combinations obtained with our approach (14) when V_{size} was set to 16, 32, and 64, respectively.

TABLE X
CORRELATION COEFFICIENTS BETWEEN THE FEATURES AND THE OPTIMALLY FITTED PARAMETERS.

Point Cloud	p_1	p_2	p_3	$CFGD$	$CBMV$	$G_{\delta_{cur}}$	C_M	C_E
bag	0.22319	0.18284	6.34197	12.20590	34.50080	0.00133	0.43300	7.30886
$banana$	0.24698	0.07960	23.60120	4.03620	9.96753	0.00140	0.85501	5.92882
$biscuits$	0.14259	0.15605	12.07180	4.51340	22.48190	0.00103	0.71598	7.17266
$cake$	0.24050	0.12525	10.48890	3.69012	22.7663	0.00100	0.36926	7.54163
$cauliflower$	0.24602	0.17673	9.77260	2.76597	11.52040	0.00049	0.75158	6.76385
$flowerpot$	0.29102	0.07488	16.21250	2.22421	22.28120	0.00045	0.44219	7.35376
$house$	0.21954	0.26889	3.59743	6.12132	28.42960	0.00054	0.52203	7.49936
$litchi$	0.19542	0.26622	3.87388	4.42747	16.50710	0.00047	0.41645	6.84626
$mushroom$	0.16402	0.22484	18.57930	6.18590	21.18330	0.00081	0.44753	7.43401
$ping-pong\_bat$	0.23975	0.22084	14.24030	3.89763	15.04600	0.00073	0.23927	6.30180
$puer\_tea$	0.12360	0.29676	11.92060	5.45375	18.04140	0.00055	0.52782	6.64002
$pumpkin$	0.13105	0.22342	7.42399	5.32005	20.88960	0.00042	0.49024	6.85955
$ship$	0.26836	0.06817	16.75630	3.56163	25.5532	0.00066	0.53502	7.43531
$statue$	0.25405	0.14214	18.77700	3.47480	21.93990	0.00057	0.66794	7.13619
$stone$	0.17022	0.29072	4.55470	9.92772	24.1395	0.00055	0.57191	6.98784
$tool\_box$	0.11718	0.26650	15.15200	3.34906	12.22590	0.00067	0.22746	5.62698
CC_{p_1}				**0.40255**	0.12046	0.11373	0.17660	0.31633
CC_{p_2}				**0.49860**	0.03450	0.38880	0.35940	0.22350
CC_{p_3}				**0.49860**	**0.41670**	0.32199	0.22125	0.29720



(a) (b)

Fig. 12. Reliability of the proposed method. The (a) PLCC and (b) SRCC between the ground truth MOS and predicted MOS are computed for 1000 random combinations of the point cloud, Q_g, and Q_c.

TABLE XI
CORRELATION COEFFICIENT BETWEEN THE PREDICTED MODEL PARAMETERS AND THE OPTIMALLY FITTED MODEL PARAMETERS ($CFGD$, $CBMV$, $G_{\delta_{cur}}$, C_M, AND C_E ARE DENOTED BY ①, ② ③, ④, AND ⑤, RESPECTIVELY. CC_n DENOTES THE CC OF THE n-TH ($n \in \{1,2,3\}$) REPEATABILITY EXPERIMENT WHILE CC_{ave} INDICATES THE AVERAGE CC OF THE EXPERIMENTS).

Number of Features	Feature Index	CC_1	CC_2	CC_3	CC_{ave}
2	①, ②	0.9161	0.8687	0.8622	0.8823
3	①, ②, ③	0.9116	0.8563	0.9242	0.8973
	①, ②, ④	0.8504	0.8611	0.8194	0.8437
	①, ②, ⑤	0.8847	0.8504	0.8748	0.8699
4	①, ②, ③, ④	0.8941	0.8552	0.8977	0.8823
	①, ②, ③, ⑤	0.9082	0.8492	0.8831	0.8879
	①, ②, ④, ⑤	0.6117	0.5793	0.2666	0.7565
5	①, ②, ③, ④, ⑤	0.6621	0.8041	0.8915	0.7859

On the other hand, QP_G^F and QP_C^F are the geometry and color QP obtained with exhaustive search by maximizing the actual MOS under the given bitrate constraint. The QP error $\Delta QP^M = |QP_G^M - QP_G^F| + |QP_C^M - QP_C^F|$ was used to measure the performance. We can see that the performance of the proposed quality model for rate control is satisfactory for various values of the CBMV block size (8, 16, 32, and 64). All the average ΔQP^Ms for block sizes 8, 16, 32, and 64 were equal to 6 which corresponds to one search step size (i.e., 6) of the exhaustive search algorithm. We can also conclude that the CBMV block size has little effect on the performance of the rate control algorithm.

VIII. CONCLUSION

We proposed an RR linear model that accurately predicts the perceptual quality of V-PCC compressed point clouds from the geometry and color quantization step sizes. The three coefficients of our model are estimated using a training set of reference point clouds and two features (CFGD and CBMV) extracted from the test reference point cloud. Results on the WPC dataset show that the PLCC and the SRCC between the

TABLE XII
INFLUENCE OF CBMV BLOCK SIZE ON RATE CONTROL PERFORMANCE.

Point cloud	R_T	QP_G^F	QP_C^F	QP_G^{M8}	QP_C^{M8}	ΔQP^{M8}	QP_G^{M16}	QP_C^{M16}	ΔQP^{M16}	QP_G^{M32}	QP_C^{M32}	ΔQP^{M32}	QP_G^{M64}	QP_C^{M64}	ΔQP^{M64}
bag	170	32	50	50	50	18	50	50	18	50	50	18	50	50	18
	510	26	44	32	44	6	32	44	6	32	44	6	32	44	6
	1495	26	38	26	32	6	26	32	6	26	32	6	26	32	6
	2130	32	32	26	26	12	26	26	12	26	26	12	26	26	12
banana	40	50	50	44	44	12	44	44	12	44	44	12	44	44	12
	120	32	44	32	38	6	32	38	6	32	38	6	32	38	6
	310	32	32	26	32	6	26	32	6	26	32	6	26	32	6
	850	26	32	26	26	6	26	26	6	26	26	6	26	26	6
cake	110	38	44	**38**	**44**	**0**	**38**	**44**	**0**	**38**	**44**	**0**	**32**	**44**	**6**
	170	32	44	32	38	6	32	38	6	32	38	6	32	38	6
	265	26	38	26	38	0	26	38	0	26	38	0	26	38	0
	460	26	38	26	32	6	26	32	6	26	32	6	26	32	6
flowerpot	75	32	44	**32**	**38**	**6**	**32**	**44**	**0**	**32**	**44**	**0**	**32**	**44**	**0**
	135	26	38	26	38	0	26	38	0	26	38	0	26	38	0
	265	26	32	26	32	0	26	32	0	26	32	0	26	32	0
	405	26	32	26	26	6	26	26	6	26	26	6	26	26	6
litchi	110	44	44	38	44	6	38	44	6	38	44	6	38	44	6
	250	38	38	32	38	6	32	38	6	32	38	6	32	38	6
	565	32	32	26	32	6	26	32	6	26	32	6	26	32	6
	1200	26	32	26	26	6	26	26	6	26	26	6	26	26	6
mushroom	50	50	50	44	50	6	44	50	6	44	50	6	44	50	6
	150	38	44	32	44	6	32	44	6	32	44	6	32	44	6
	220	38	44	32	38	12	32	38	12	32	38	12	32	38	12
	375	26	38	26	38	0	26	38	0	26	38	0	26	38	0
puer_tea	75	44	50	32	50	12	32	50	12	32	50	12	32	50	12
	190	26	44	26	38	6	26	38	6	26	38	6	26	38	6
	640	26	32	26	32	0	26	32	0	26	32	0	26	32	0
	1525	26	26	26	26	0	26	26	0	26	26	0	26	26	0
statue	55	50	50	44	50	6	44	50	6	44	50	6	44	50	6
	105	32	50	38	44	12	38	44	12	38	44	12	38	44	12
	155	32	44	32	44	0	32	44	0	32	44	0	32	44	0
	200	32	44	32	38	6	32	38	6	32	38	6	32	38	6
Average				-	-	**6**	-	-	**6**	-	-	**6**	-	-	**6**

predicted MOSs and the actual MOSs are about 0.91, which indicates that the proposed model is highly accurate.

Moreover, we showed how the proposed model can be used to optimize the perceptual quality subject to a target bitrate. Due to the accuracy of the proposed model, the subjective quality of the point clouds encoded with the proposed algorithm is much better than that of $P2P_{ES}$.

As future work, we will assess the performance of the proposed model on the newly provided high quality point clouds of the MPEG PCC group. We will also apply the proposed quality metric to quality enhancement for point clouds.

REFERENCES

[1] S. Gu, J. Hou, H. Zeng, H. Yuan, and K.-K. Ma, "3D point cloud attribute compression using geometry-guided sparse representation," *IEEE Transactions on Image Processing*, vol. 29, pp. 796–808, 2019.

[2] L. Li, Z. Li, V. Zakharchenko, J. Chen, and H. Li, "Advanced 3D motion prediction for video-based dynamic point cloud compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 289–302, 2020.

[3] S. Gu, J. Hou, H. Zeng, and H. Yuan, "3D point cloud attribute compression via graph prediction," *IEEE Signal Processing Letters*, vol. 27, pp. 176–180, 2020.

[4] 3DG, "Text of ISO/IEC CD 23090-9 geometry-based point cloud compression," *Doc. ISO/IEC JTC1/SC29/WG11 MPEG N18478, Geneeve, Switzerland*, Mar. 2019.

[5] 3DG, "Text of ISO/IEC CD 23090-5: Video-based point cloud compression," *ISO/IEC JTC1/SC29/WG11 MPEG N18030, Macau, China*, Oct. 2018.

[6] H. Liu, H. Yuan, Q. Liu, J. Hou, and J. Liu, "A comprehensive study and comparison of core technologies for MPEG 3-D point cloud compression," *IEEE Transactions on Broadcasting*, vol. 66, no. 3, pp. 701–717, Dec.2020.

[7] D. Girardeau-Montaut, M. Roux, R. Marc, and G. Thibault, "Change detection on points cloud data acquired with a ground laser scanner," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 36, p. W19, 2005.

[8] D. Tian, H. Ochimizu, C. Feng, R. Cohen, and A. Vetro, "Geometric distortion metrics for point cloud compression," in *2017 IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3460–3464.

[9] R. Mekuria, Z. Li, C. Tulvan, and P. Chou, "Evaluation criteria for PCC (point cloud compression)," *ISO/IEC JTC1/SC29/WG11 MPEG N16332*, 2016.

[10] E. Alexiou and T. Ebrahimi, "Point cloud quality assessment metric based on angular similarity," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, 2018, pp. 1–6.

[11] G. Meynet, J. Digne, and G. Lavoué, "PC-MSDM: A quality metric for 3D point clouds," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*, 2019, pp. 1–3.

[12] E. Alexiou, T. Ebrahimi, M. V. Bernardo, M. Pereira, A. Pinheiro, L. A. D. S. Cruz, C. Duarte, L. G. Dmitrovic, E. Dumic, D. Matkovics *et al.*, "Point cloud subjective evaluation methodology based on 2D rendering," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1–6.

[13] E. Alexiou and T. Ebrahimi, "On subjective and objective quality evaluation of point cloud geometry," in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1–3.

[14] Z. Ma, M. Xu, Y.-F. Ou, and Y. Wang, "Modeling of rate and perceptual quality of compressed video as functions of frame rate and quantization stepsize and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 5, pp. 671–682, 2011.

[15] Z. Ma, M. Xu, K. Yang, and Y. Wang, "Modeling of rate and perceptual quality of video and its application to frame rate adaptive rate control," in *2011 18th IEEE International Conference on Image Processing*, 2011,

pp. 3321–3324.
[16] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, "Subjective and objective quality evaluation of compressed point clouds," in *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, Oct 2017, pp. 1–6.
[17] E. Alexiou and T. Ebrahimi, "On the performance of metrics to predict quality in point cloud representations," in *Applications of Digital Image Processing XL*, vol. 10396, 2017.
[18] E. Alexiou and T. Ebrahimi, "Impact of visualisation strategy for subjective quality assessment of point clouds," in *2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2018, pp. 1–6.
[19] E. M. Torlig, E. Alexiou, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A novel methodology for quality assessment of voxelized point clouds," in *Applications of Digital Image Processing XLI*, vol. 10752, 2018.
[20] Q. Yang, H. Chen, Z. Ma, Y. Xu, R. Tang, and J. Sun, "Predicting the perceptual quality of point cloud: A 3D-to-2D projection-based exploration," *IEEE Transactions on Multimedia*, 2020, doi: 10.1109/TMM.2020.3033117.
[21] H. Su, Z. Duanmu, W. Liu, Q. Liu, and Z. Wang, "Perceptual quality assessment of 3D point clouds," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 3182–3186.
[22] A. Javaheri, C. Brites, F. M. B. Pereira, and J. M. Ascenso, "Point cloud rendering after coding: Impacts on subjective and objective quality," *IEEE Transactions on Multimedia*, 2020, doi: 10.1109/TMM.2020.3037481.
[23] E. Alexiou, I. Viola, T. M. Borges, T. A. Fonseca, R. L. de Queiroz, and T. Ebrahimi, "A comprehensive study of the rate-distortion performance in MPEG point cloud compression," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019.
[24] E. Zerman, P. Gao, C. Ozcinar, and A. Smolic, "Subjective and objective quality assessment for volumetric video compression," *Electronic Imaging*, vol. 2019, no. 10, pp. 323–1–323–7(7), 2019.
[25] Q. Yang, Z. Ma, Y. Xu, Z. Li, and J. Sun, "Inferring point cloud quality via graph similarity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, doi: 10.1109/TPAMI.2020.3047083.
[26] E. Alexiou and T. Ebrahimi, "Towards a point cloud structural similarity metric," in *2020 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2020.
[27] G. Meynet, Y. Nehmé, J. Digne, and G. Lavoué, "PCQM: A full-reference quality metric for colored 3D point clouds," in *12th International Conference on Quality of Multimedia Experience (QoMEX 2020)*, 2020.
[28] I. Viola and P. Cesar, "A reduced reference metric for visual quality evaluation of point cloud contents," *IEEE Signal Processing Letters*, vol. 27, pp. 1660–1664, 2020.
[29] Q. Liu, H. Yuan, J. Hou, H. Liu, and R. Hamzaoui, "Model-based encoding parameter optimization for 3D point cloud compression," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 2018, pp. 1981–1986.
[30] Q. Liu, H. Yuan, J. Hou, R. Hamzaoui, and H. Su, "Model-based joint bit allocation between geometry and color for video-based 3D point cloud compression," *IEEE Transactions on Multimedia*, 2020 DOI:10.1109/TMM.2020.3023294.
[31] Q. Liu, H. Yuan, R. Hamzaoui, and H. Su, "Coarse to fine rate control for region-based 3D point cloud compression," in *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. IEEE, 2020, pp. 1–6.
[32] H. Yuan, Y. Chang, J. Huo, F. Yang, and Z. Lu, "Model-based joint bit allocation between texture videos and depth maps for 3-D video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 4, pp. 485–497, 2011.
[33] H. Yuan, S. Kwong, C. Ge, X. Wang, and Y. Zhang, "Interview rate distortion analysis-based coarse to fine bit allocation algorithm for 3-D video coding," *IEEE transactions on broadcasting*, vol. 60, no. 4, pp. 614–625, 2014.
[34] MPEG 3DG. (2019) V-PCC test model v7. [Online]. Available: http://mpegx.int-evry.fr/software/MPEG/PCC/TM/mpeg-pcc-tmc2.git
[35] Y. Nehmé, J.-P. Farrugia, F. Dupont, P. LeCallet, and G. Lavoué, "Comparison of subjective methods, with and without explicit reference, for quality assessment of 3D graphics," in *ACM Symposium on Applied Perception 2019*, 2019, pp. 1–9.
[36] A. M. Van Dijk, J.-B. Martens, and A. B. Watson, "Quality asessment of coded images using numerical category scaling," in *Advanced Image and Video Communications and Storage Technologies*, vol. 2451, 1995, pp. 90–101.
[37] ITU, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation BT.500-13*, 2012.
[38] E. R. Girden, *ANOVA: Repeated measures*. Sage Publications, Inc, 1992.
[39] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Transactions on image processing*, vol. 4, no. 6, pp. 713–724, 1995.
[40] R. Mekuria, S. Laserre, and C. Tulvan, "Performance assessment of point cloud compression," in *2017 IEEE Visual Communications and Image Processing (VCIP)*, 2017, pp. 1–4.
[41] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, and G. Zhai, "No-reference quality assessment of contrast-distorted images based on natural scene statistics," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 838–842, 2014.
[42] S. Rimac-Drlje, D. Zagar, and G. Martinovic, "Spatial masking and perceived video quality in multimedia applications," in *2009 16th International Conference on Systems, Signals and Image Processing*, 2009, pp. 1–4.
[43] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2013.
[44] D. Thanou, P. A. Chou, and P. Frossard, "Graph-based compression of dynamic 3D point cloud sequences," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1765–1778, 2016.
[45] A. J. Dobson and A. G. Barnett, *An introduction to generalized linear models*. CRC press, 2018.
[46] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Transactions on image processing*, vol. 15, no. 2, pp. 430–444, 2006.
[47] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
[48] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, 2003, pp. 1398–1402.
[49] S. E. Palmer, *Vision science: Photons to phenomenology*. MIT press, 1999.
[50] Y. Dai. (2017) Compute underlying surface's normal and curvature in local neighborhoods. [Online]. Available: https://www.mathworks.com/matlabcentral/mlc-downloads/downloads/submissions/61125/versions/3/previews/NormalAtEachPoint/normal.m/index.html
[51] ITU, "Parameter values for the hdtv standards for production and international programme exchange," *Recommendation BT.709-6*, 2015.
[52] I. Viola and P. Cesar. (2020) PCM_RR. [Online]. Available: https://github.com/cwi-dis/PCM_RR

Qi Liu received the B.S. degree from Shandong Technology and Business University, Shandong, China, in 2011 and the M.S. degree from the School of Telecommunication Engineering, Xidian University, Xi'an, China, in 2014. Currently, she is pursuing the Ph.D degree with the Information Science and Engineering from Shandong University, Shandong, China. From 2018.09-2019.08, she also worked as a visiting graduate student with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. Her research interests include point cloud coding, processing and quality assessment, etc.

Hui Yuan (S'08-M'12-SM'17) received the B.E. and Ph.D. degree in telecommunication engineering from Xidian University, Xi'an, China, in 2006 and 2011, respectively. From 2011.04 to now, he works as a lecturer (2011.04-2014.12), an associate Professor (2015.01-2016.08), and a full professor (2016.09-), at Shandong University (SDU), Jinan, China. From 2013.01-2014.12, and 2017.11-2018.02, he also worked as a post-doctor fellow (Granted by the Hong Kong Scholar Project) and a research fellow, respectively, with the department of computer science, City University of Hong Kong (CityU). His current research interests include video/image/immersive media processing, compression, adaptive streaming, and computer vision, etc.

He served as an Area Chair of IEEE ICME 2021, IEEE ICME 2020, and IEEE VCIP 2020.

Raouf Hamzaoui (M02CSM07) received the M.Sc. degree in mathematics from University of Montreal, Montral, QC, Canada, in 1993; the Dr.rer.nat. degree from University of Freiburg, Freiburg im Breisgau, Germany, in 1997; and the Habilitation degree in computer science from University of Konstanz, Konstanz, Germany, in 2004. He was an Assistant Professor with the Department of Computer Science, University of Leipzig, Leipzig, Germany, and with the Department of Computer and Information Science, University of Konstanz. He joined De Montfort University, Leicester, U.K., in 2006, where he is currently a Professor in Media Technology. His research interests include image and video coding, multimedia communication systems, error control systems, and machine learning. He served as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2010-2016) and the IEEE TRANSACTIONS ON MULTIMEDIA (2017-2021).

Honglei Su (M'19) received the B.A.Sc degree from Shandong University of Science and Technology, Qingdao, China in 2008 and the Ph.D. degree from Xidian University, Xi'an, China in 2014. From 2014.09 to now, he works as an Assistant Professor with the School of Electronic Information, Qingdao University, Qingdao, China. From 2018.03-2019.03, he also worked as a visiting scholar with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests include perceptual image processing, immersive media processing, computer vision, etc.

Junhui Hou (S'13-M'16-SM'20) received the B.Eng. degree in information engineering (Talented Students Program) from the South China University of Technology, Guangzhou, China, in 2009, the M.Eng. degree in signal and information processing from Northwestern Polytechnical University, Xian, China, in 2012, and the Ph.D. degree in electrical and electronic engineering from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, in 2016. He immediately joined the Department of Computer Science, City University of Hong Kong, as an Assistant Professor in Jan. 2017. His research interests fall into the general areas of visual computing, such as image/video/3D geometry data representation, processing and analysis, semi/un-supervised data modeling, and data compression and adaptive transmission.

Dr. Hou was the recipient of several prestigious awards, including the Chinese Government Award for Outstanding Students Study Abroad from China Scholarship Council in 2015 and the Early Career Award (3/381) from the Hong Kong Research Grants Council in 2018. He is an elected member of MSA-TC and VSPC-TC, IEEE CAS. He is currently serving as an Associate Editor for IEEE Transactions on Circuits and Systems for Video Technology and Signal Processing: Image Communication. He also served as an Associate Editor for The Visual Computer, the Guest Editor for the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, and an Area Chair of ACM MM 2019-2021, IEEE ICME 2020, and WACV 2021. He is a senior member of IEEE.

Huan Yang received the Ph.D. degree in computer engineering from Nanyang Technological University, Singapore, in 2015, the M.S. degree in computer science from Shandong University, China, in 2010, and the B.S. degree in computer science from the Heilongjiang Institute of Technology, China, in 2007. Currently, she is working in the College of Computer Science and Technology, Qingdao University, Qingdao China. Her research interests include image/video processing and analysis, perception-based modeling and quality assessment, object detection/recognition, and machine learning.